STEREO PERFORMANCE FOR CLUTTERED SCENES

Mannan F., Langer M. S. - McGill University
{fmannan, langer}@cim.mcgill.ca

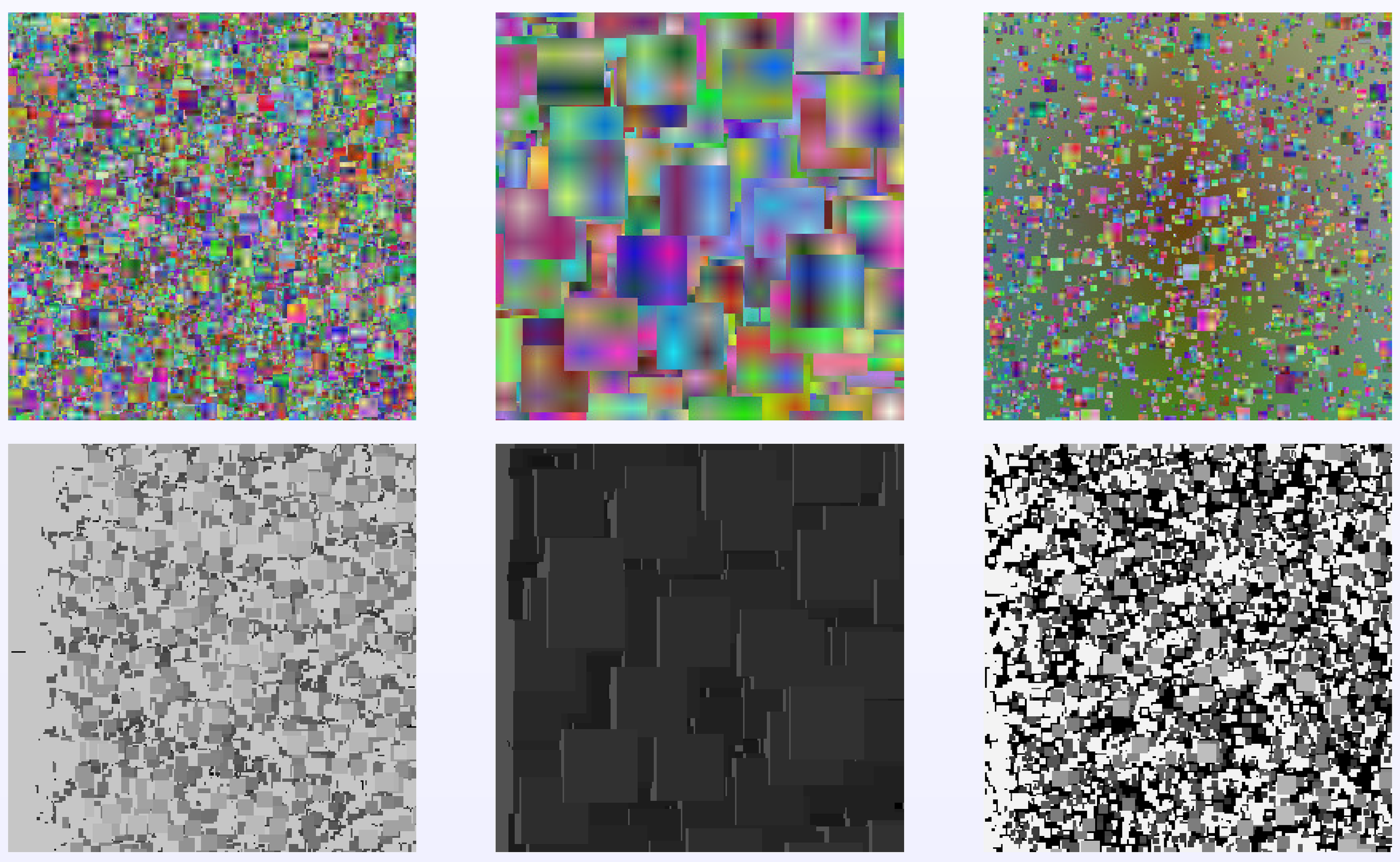
McGill
Centre for Intelligent Machines

Abstract

Previous works on stereo evaluation did not address performance for different types of scenes. This work [1], evaluates MRF-based stereo formulations for cluttered scenes [2]. Three types of methods are considered: basic (**Basic**)[3], uniqueness (**KZ-Uni**)[4], and visibility (**KZ-vis**)[5]. These are evaluated based on mislabeled pixels of different types (binocular/monocular) in different regions (on or away from occlusion boundary).

What are cluttered scenes?

These are scenes with large number of objects of certain size distribution uniform randomly distributed in 3D space. They have significantly more depth discontinuity and occlusion than general scenes. Shrubs, bushes, tree canopies etc. are some examples of natural cluttered scene. Sample synthetic scenes with disparity and occlusion (brightest) are shown below.



Key Observations

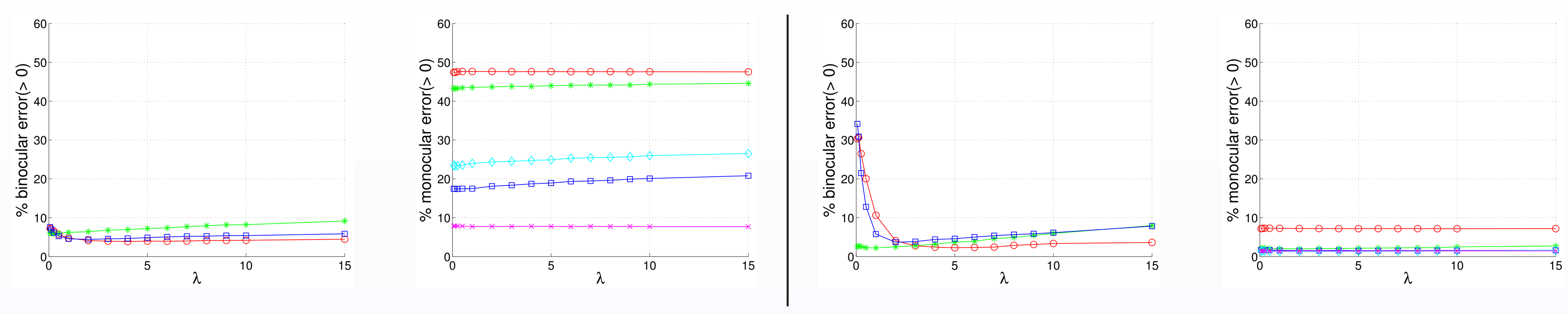
- Binocular pixels near occlusion boundary makes significant contribution to the overall accuracy.
- **KZ-vis** does best for monocular pixels.
- **Basic** does best for binocular pixels.
- **Basic-cc** does best for occlusion labeling.

Stereo Formulations

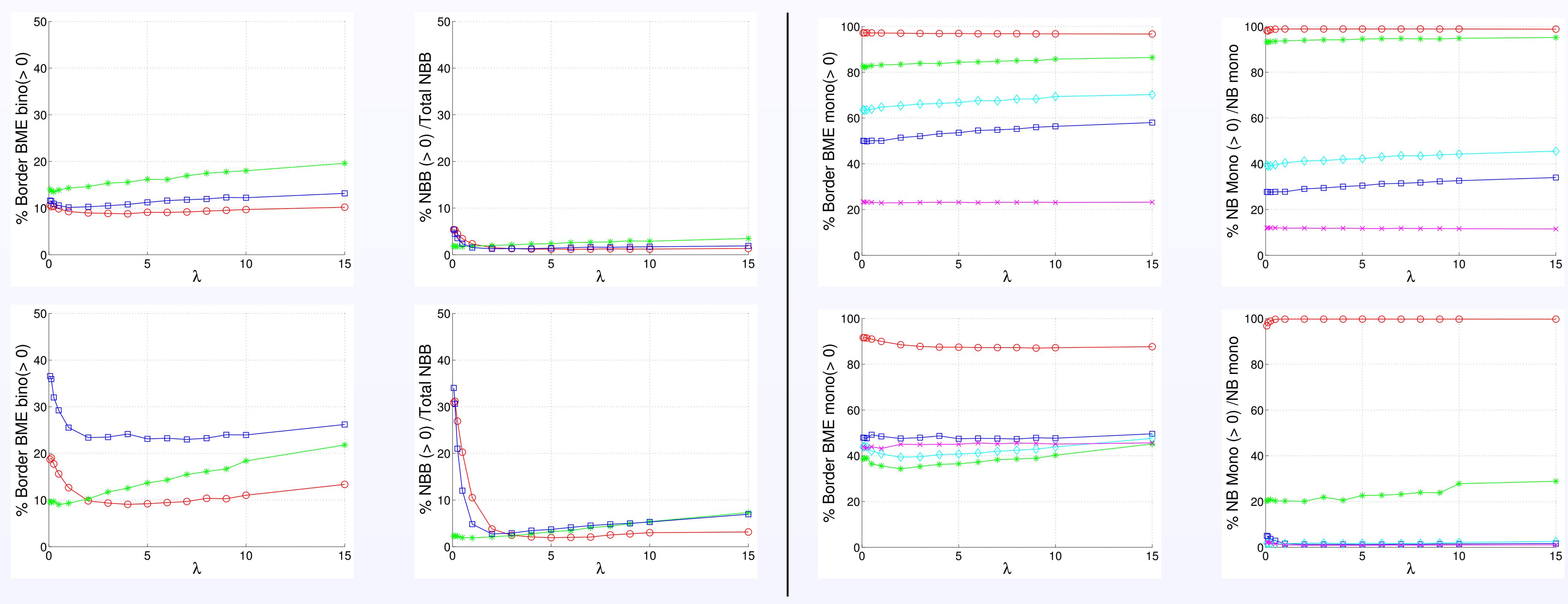
Basic: $E = \sum_p d(I_l, I_r, p, q)^2 + \lambda \sum_{p,q} w_{pq}\, min(|f_p - f_q|, \tau)$
KZ-Uni: Pairs of pixels are considered as nodes in the random field. Labels are binary indicating whether they correspond or not.
KZ-vis: Maintains visibility constraint in addition to the basic constraints in every iteration. The constraint ensures that a labeling does not block visibility of a pixel in the other view.
Occlusion Labeling: **Basic** and **KZ-vis** do not have any occlusion labeling. With cross-checked occlusion labeling they are refered to as **Basic-cc** and **KZ-vis-cc**.

Experiments

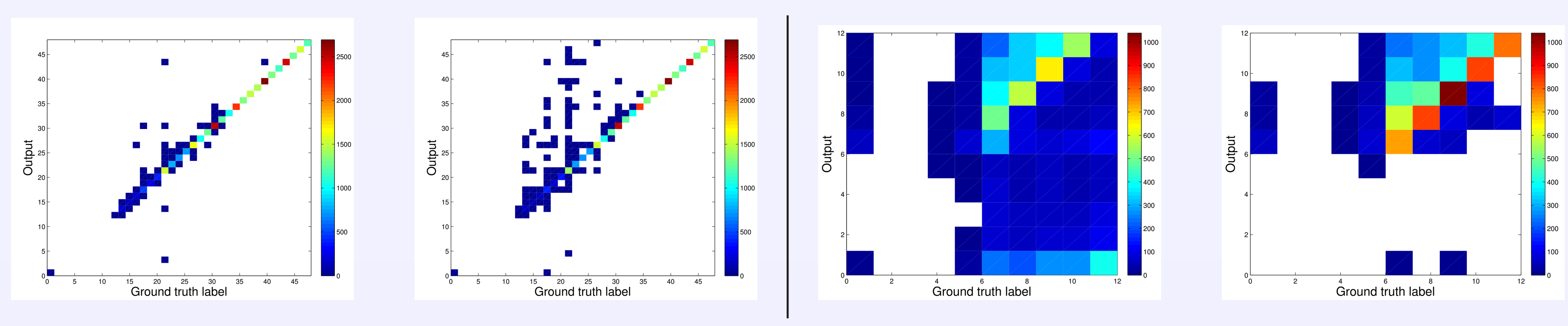
The following figure shows the error for binocular and monocular pixels for two different occlusion amount (left two dense and right two sparse).



The following figure shows error for pixels on the binocular-monocular boundary (Border BME) and away from the boundary (NBB) for both binocular (left half) and monocular (right half) pixels. The first row is for a dense and the second row for a sparse scene.



The following figure shows joint ground-truth and disparity labeling for binocular (left half) and monocular (right half) pixels for densely occluded scenes. The first of each set (i.e. 1st and 3rd col.) corresponds to **Basic** and the second (i.e. 2nd and 4th col.) to **KZ-vis**. Diagonal entries correspond to correct labeling while the off-diagonals to incorrect labeling.



References

[1] F. Mannan and M.S. Langer, "Performance of Stereo Methods in Cluttered Scenes," in *Computer and Robot Visiion*, 2011.
[2] M. S. Langer, "Surface visibility probabilities in 3d cluttered scenes," in *ECCV '08: Proceedings of the 10th ECCV*. Springer-Verlag, 2008, pp. 401–412.
[3] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *TPAMI*, vol. 30, no. 6, pp. 1068–1080, June 2008.
[4] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions via graph cuts," in *ICCV*, 2001, pp. 508–515.
[5] ——, "Multi-camera scene reconstruction via graph cuts," in *ECCV (3)*, ser. Lec. Notes in Comp. Sci., vol. 2352. Springer, 2002, pp. 82–96.